

Building pattern recognition applications with the SPARE library

Lorenzo Livi^{*†1}, Guido Del Vescovo^{‡2}, Antonello Rizzi^{§2}, and Fabio Massimo Frattale Mascioli^{¶2}

¹Dept. of Computer Science, Ryerson University, 350 Victoria Street, Toronto, ON M5B 2K3, Canada

²Dept. of Information Engineering, Electronics, and Telecommunications, SAPIENZA University of Rome, Via Eudossiana 18, 00184 Rome, Italy

February 23, 2015

Abstract

This paper presents the SPARE C++ library, an open source software tool conceived to build pattern recognition and soft computing systems. The library follows the requirement of the generality: most of the implemented algorithms are able to process user-defined input data types transparently, such as labeled graphs and sequences of objects, as well as standard numeric vectors. Here we present a high-level picture of the SPARE library characteristics, focusing instead on the specific practical possibility of constructing pattern recognition systems for different input data types. In particular, as a proof of concept, we discuss two application instances involving clustering of real-valued multidimensional sequences and classification of labeled graphs.

Index terms— Software library for pattern recognition; Structured data processing; C++ programming; Templates.

1 Introduction

The classic pattern recognition discipline is essentially conceived for the \mathbb{R}^n vector space geometry. However, many interesting applications, coming for instance from computational biology, multimedia intelligent processing, and computer vision, deal with structured patterns, such as images, audio and video sequences, strings of characters, and labeled graphs [11, 16, 49]. Usually, in order to take advantage of the existing data-driven modeling systems, each pattern of a structured domain \mathcal{X} is reduced to a set of real-valued features by adopting a preprocessing function, tailored on the specific application at hand [9, 25, 28, 30, 37, 38, 40]. The design of this preprocessing procedure is a critical task, since useful information could be lost due to an excessive information compression. For this reason, it could be useful to design pattern recognition systems able to deal directly with structured or unconventional input domains. Consequently, the availability of effective models and algorithms able to represent and process a set of samples belonging to \mathcal{X} is fundamental [3, 14, 17, 19]. When dealing with cluster analysis, the abstraction of a cluster of objects by a representative model is useful for two main reasons. The first reason is about computational issues. For example, having to compare a particular pattern with the representative instead of each single element in the cluster often results in a significative speed up. The second reason deals with the ability of a learning system to create

^{*}llivi@scs.ryerson.ca

[†]Corresponding author

[‡]delvescovo@uniroma1.it

[§]antonello.rizzi@uniroma1.it

[¶]mascioli@infocom.uniroma1.it

abstract knowledge, a fundamental feature of the cerebral cortex in biological brains. In fact, the human ability to synthesize concepts from a set of objects plays a key role in human reasoning and thus in any machine learning procedure.

SPARE – Something for PAttern REcognition – is a C++ library conceived as an essential software framework for rapid development of applications of pattern recognition and computational intelligence. SPARE is based on the *concept-driven* template meta-programming technique [1], aiming to provide basic classes and procedures for rapid application development when dealing with clustering algorithms, inductive modeling systems (classification and function approximation), and optimization problems by neural networks, fuzzy logic, and evolutionary computation. The set of concepts and classes has been thought in order to capture the essence of some basic yet important algorithms, and reflects the fact that even the most complex data-driven modeling system can be conceived as an organized ensemble of small code blocks, each of them implementing specific functionalities with a standard interface. Designing a library with concepts in mind permits to create a set of classes that are not relegated to a fixed hierarchy: concepts are *non-restricting* design choices, which in essence relaxes the development process. The meta-algorithms implemented in SPARE can be easily used to build pattern recognition applications dealing with customizable (i.e., user-defined) data types, without any strict necessity to represent objects as feature-based patterns. Moreover, the template-based approach simplifies the interoperability with other state-of-the-art software libraries [10, 22] and with the code written by the user. Being written in standard C++, SPARE code is also efficient and fairly portable to several major operating systems and architectures. SPARE provides multi-thread implementations of different algorithms. This is an important feature of modern hardware, which should be exploited in any computationally intensive application.

SPARE is an open source software library, released under the GNU GPL 3.0 license. The main site of the project is located at <https://sourceforge.net/p/libspare/home/Spares/>, where it is possible to obtain an up-to-date version of the SPARE library, as well as all the available documentation, tutorials, and some focused benchmarks considering other well-known software libraries. We decided to release SPARE as an open source project since, in our opinion, using open source software in research projects allows to easily reproduce and verify experimental results, therefore improving the transparency of the research itself. Moreover, it allows a faster exchange of knowledge between different research groups.

In this paper, we present the high-level design and functionalities of the SPARE library together with two practical application instances. The first one concerns the clustering of a set of sequences of multidimensional real-valued vectors, and aims to analyze the modeling capability of one of the implemented cluster representative techniques in SPARE. The second one discusses how to develop and test a graph-based classification system with SPARE. A benchmark of four graph matching algorithms on a batch of 15 synthetic classification problem instances for labeled graphs is contextually analyzed. The herein reported applications are introduced only as a proof of concept to show the versatility of the SPARE library in designing pattern recognition systems for different data types. Therefore, the presented experimental results should not be considered as “research results”. Technical details regarding the library are omitted in this paper, referring the reader to the SPARE documentation. The involved theoretical and methodological aspects are not discussed as well; we provide literature references for this purpose. The remainder of the paper is structured as follows. Section 2 depicts the SPARE design, highlighting its main features. In Section 3 we introduce the concept of *generic algorithm*, showing two specific test cases; in particular in Section 3.2 we discuss an experiment regarding clustering of sequences, while in Section 3.3 we show the application of some graph matching algorithms on synthetically generated classification problems for labeled graphs. Finally, Section 4 concludes the paper pointing at the SPARE future directions.

2 The SPARE Library

2.1 A Glimpse of Template Metaprogramming

A *metaprogram* can be defined as a program that can manipulate itself as data or that can execute part of the job at *compile time*. The language in which the metaprogram is written is called the *metalanguage*, and

the capability of a language to be its own metalanguage is called *reflection* [1, 2]. Metaprograms work with an input language called *domain language* and output language called *host language*. This abstraction is provided for two main reasons: to facilitate the programmer in writing high level programs and to provide an abstraction of mechanism of program manipulation. An example of this kind of programs is the YACC parser [20], which can be referred to as a *translator* – a metaprogram whose domain language is different from the host language. *Template metaprogramming* is a popular *native language* metaprogramming technique employed by the C++ programming language, which allows to manipulate a program at compile time using *templates*. The compiler translates the template generating a temporary source code that will be successively merged with the rest of the program [1, 46, 50, 51]. Templates are an instance of the well-known *generic programming* paradigm, in which algorithms are written in terms of the so-called *to-be-specified-later* types [47]. The exact type is specified as a class parameter when the algorithm is used. This mechanism permits to write reusable algorithms that abstract the data type with which they work. Other languages, such as Haskell, D, ML, Caml, XL, and modern Java, adopt a similar technique.

For example, a well-known family of classes in the C++ STL are containers [4, 6]. A container is a *concept* that is able to store a generic type of data, providing many functionalities for manipulations, like insert, delete, and retrieval of elements. Containers are ubiquitous in every modern programming languages, such as the *Java Collections* framework [32]. The internal low level data structures used by the specific container class are hidden to the programmer, and only high level functionalities are exposed.

A C++ concept can be defined as “a specification of the requirements on a set of template arguments” [13]. Concepts are (informal) type systems for templates, aiming to offer a unified way to define function signatures. The main issues in defining a formal type system for templates using concepts are performance requirements. A concept must be rigorous enough to be able to define the requirements of a specific group of classes and should avoid exposing too many internal specifications to the programmer. For example the concept of *iterator* defines a family of classes that are able to fetch items stored in an abstract collection. The specific iteration strategy is hidden from the user point of view, and, of course, the specific data type to iterate must be defined through the template mechanism.

Many prominent C++ libraries follow this kind of specification. An example is offered by the Boost library [1, 21], a very large and well organized collection of classes that spans between many areas, such as graphs, networking, and parallel computing. SPARE is designed in a similar way to Boost. In fact, the main design effort has been devoted to the definition of a suited collection of concepts, each one conceived in terms of classes-objects providing similar functionalities.

2.2 Main SPARE Features

A list of concepts and corresponding models (i.e., algorithms) available in the latest SPARE release is shown in Tables 1 and 2. The *Dissimilarity* concept is a generalization of a distance measure, which in SPARE provides models able to process different data types. The dissimilarity measure is one of the funding blocks of many pattern recognition systems, especially when dealing with non-conventional input domains. The *Similarity* concept is practically the dual concept of the dissimilarity, and it is applicable to a broad range of data types as well. Most of the currently implemented similarity measures are well-known kernel functions widely used in pattern recognition applications. The *Evaluator* provides the abstraction of a (real-valued) function of the type $f : \mathbb{R}^d \rightarrow \mathbb{R}$; the *MultiEvaluator* is the direct extension of the Evaluator concept which just generalizes the function to multidimensional outputs, i.e., $f : \mathbb{R}^d \rightarrow \mathbb{R}^m$. By combining Evaluator and MultiEvaluator, it is possible to design well-known *fuzzy inference systems* (please refer to the SPARE technical documentation for a nice “how to” guide). The *Representative* concept provides algorithms to calculate and update the representative of a set of patterns. In general, models implementing such a concept are applicable to different data types, while others are applicable by definition only to \mathbb{R}^d data. The *Optimizer* concept includes optimization algorithms; the *Environment* concept is intimately related to Optimizer since it provides specific solutions to model the so-called *individuals* (i.e., the candidate solutions) in the implemented genetic algorithm. The *Representation* concept collects algorithms that aim to produce a new matrix-based representation of the input data (usually it is applicable to any input data type by means of a suited user-defined dissimilarity measure). Such a new representation could be used, for instance, to address the original

problem in a suited *embedding space*. *Clustering* is an auto-explicative concept collecting of course clustering algorithms, which are applicable, according to their current implementations, to virtually any input data type for which it is possible to define a dissimilarity measure. The *Supervised* concept accounts for classification systems that are trained in a supervised fashion, i.e., they are trained by making use of a labeled training set. Finally, the *Unsupervised* concept provides instead classifiers which are trained in an unsupervised fashion (e.g., by means of a clustering-based model synthesis).

Table 3 provides the functionalities currently available to manage data represented as labeled graphs. Those algorithm are presented in a separated table since they operate specifically for labeled graph data types; however, they are fully interoperable with all the other SPARE classes. Part of the herein presented concepts (see for instance Dissimilarity and Similarity) are just the specialization of the already introduced concepts for the labeled graphs data type. *MatrixRepresentation* provides models to represent an input graph as a matrix (such as the adjacency matrix or the so-called transition matrix). The *Operator* provides algorithms that implement some well-known (binary) operators among graphs. Finally, the *Seriation* concept collects algorithmic solutions to transform an input graph into a sequence of its vertex labels.

More detailed and up-to-date technical information about the library can be retrieved from the main web site: <https://sourceforge.net/p/libspare/home/Spare/>.

Table 1: List of the main SPARE concepts and implemented models – 1/2.

Concept	Implementation	Short Description
Dissimilarity	Hamming	The standard Hamming distance for $\{0, 1\}$ data vectors.
	Euclidean	The standard Euclidean distance for real valued vectors.
	Minkowski	The order p Minkowski distance for real valued vectors.
	Dtw	The Dynamic Time Warping dissimilarity between generic data sequences.
	Levenshtein	The Levenshtein edit distance.
	CBMF	The Container-based Best Match First (CBMF) dissimilarity
	Constant	The constant user-defined dissimilarity.
	Delta	The Delta dissimilarity measure ($d(x, y) = 1 \Leftrightarrow x \neq y$)
	DissimilarityMatrix	The dissimilarity matrix is an indexed matrix (Boost matrix) of pre-calculated dissimilarity values. It is possible to make access to the dissimilarity values according to the user-defined indices
Similarity	ModuleDistance	The absolute distance (also called module) of two input objects that implement the $-$ operator
	Cosine	The cosine similarity function between real-valued vectors. The function effectively computes the cosine of the angle among the two input vectors.
	ExponentialKernel	The exponential kernel function. It is defined in terms of a (symmetric) dissimilarity measure, which is assumed to be able to operate on the user-defined input space.
	GTSKernel	The Generalized T-Student kernel function. It is defined in terms of a (symmetric) dissimilarity measure, which is assumed to be able to operate on the user-defined input space.
	HyperbolicTangentKernel	The Hyperbolic tangent kernel function between real-valued vectors.
	LaplacianKernel	The Laplacian kernel function. It is defined in terms of a (symmetric) dissimilarity measure, which is assumed to be able to operate on the user-defined input space.
	PolynomialKernel	The polynomial kernel function that operates between real-valued vectors.
	RationalQuadraticKernel	The rational quadratic kernel function. It is defined in terms of a (symmetric) dissimilarity measure, which is assumed to be able to operate on the user-defined input space.
	RBFKernel	The RBF Gaussian kernel function. It is defined in terms of a (symmetric) dissimilarity measure, which is assumed to be able to operate on the user-defined input space.
Evaluator	PiecewiseLinear	A piecewise linear membership function.
	Gaussian	A Monodimensional Gaussian function.
	MultiGaussian	A Multivariate Gaussian function.
MultiEvaluator	FuzzyAntecedent	A fuzzy rule antecedent based on scalar fuzzy membership functions.
	MultiGaussianAntecedent	A fuzzy rule antecedent based on multivariate Gaussian membership functions.
	FuzzyConsequent	A fuzzy rule consequent based on scalar fuzzy membership functions.
	HyperplaneConsequent	A fuzzy rule consequent computed as a linear polynomial function of rule's inputs.
	FuzzyRuleBlock	A block of fuzzy logic rules.
Representative	Centroid	Mean vector based representative of a cluster of real valued vectors. The point to cluster dissimilarity is user defined.
	Mahalanobis	Representative based on mean vector and covariance matrix for a cluster of real valued vectors. The point to cluster dissimilarity is computed as the Mahalanobis distance.
	MinSod	Representative of a generic cluster of objects, based on the cluster sample which minimizes the Sum Of Distances from the other samples. The point to cluster dissimilarity is user defined.
	FuzzyMinSod	Same as MinSod, but it offers the functionality of computing the membership function of the considered objects. The membership function is calculated according to an Evaluator function that elaborates the distance with respect to the current MinSOD element.
	PFuzzyMinSod	Same as FuzzyMinSod, but is characterized by a different replacement policy. The elements are replaced according to a probability distribution computed considering the dissimilarity values with respect to the MinSOD element. Elements with lower dissimilarity values have lower probability of being replaced.
	Rlse	Representative of a cluster of input-output pairs, where inputs are real valued vectors and output are scalar real values, based on a hyperplane approximating the points in the cluster. The point to cluster dissimilarity is computed as the hyperplane approximation error.
	FuzzyHyperbox	Model of a fuzzy set-based hyperbox.

Table 2: (Continuation of Table 1) List of the main SPARE concepts and implemented models – 2/2.

Concept	Implementation	Short Description
Optimizer	GeneticAlgorithm	An implementation of a classic genetic algorithm.
	PGeneticAlgorithm	Same genetic algorithm implementation but with multi-threading capability.
Environment	DiscreteCode	An environment agent to be used with the genetic algorithm implementation, where the candidate solution (individual) is coded as an integer-valued vector. Classical crossover and mutation operators are provided within the class.
Representation	DissimilarityRepr	The dissimilarity representation of an input set with respect to a representation set. It is generalized according to a specific type of dissimilarity measure.
Clustering	Kmeans	The standard k -means algorithm generalized for what concerns the cluster representatives initialization scheme (we provide different initialization algorithms that can be also user-defined).
	Bsas	The Basic Sequential Algorithmic Scheme clustering algorithm. Fast linear-time cluster generation rule.
	MTBsas	Same as Bsas but with multi-threading capability.
Supervised	KnnApprox	A version of the k nearest neighborhood algorithm for function approximation problems.
	KnnClass	The well-known k nearest neighborhood algorithm for classification problems.
	MTKnnClass	Same as KnnClass but with multi-threading capability.
	SvmApprox	A version of the SVM algorithm for function approximation problems (uses libsvm as core engine).
	SvmClass	The well-known SVM algorithm for classification problems (uses libsvm as core engine).
	Scbc	A supervised clustering-based classification system for general user-defined data types.
	MMNet	A neurofuzzy Min-Max network trained with the ARC/PARC algorithms [39].
	DlibSvmClass	A wrapper of the Dlib implementation of the kernelized C-SVM and ν -SVM. It accepts a user-defined kernel function which can be generalized to virtually any data type.
	DlibSvmApprox	A wrapper of the Dlib implementation of the kernelized ϵ -SVR. It accepts a user-defined kernel function, which can be generalized to virtually any data type.
Unsupervised	Rlrpa	A reinforcement learning based classification algorithm for user-defined data types.
	Ucbc	The same as Scbc but synthesizes the classification model in an unsupervised fashion.

Table 3: List of SPARE concepts and currently implemented models for graph-based pattern analysis.

Concept	Implementation	Short Description
Dissimilarity	PGED	The Perfect GED class implements the GED algorithm that solves the GED problem optimally.
	BMF	The weighted (vertex) Best Match First graph dissimilarity measure implements a vertex assignment solution following a greedy heuristic.
	sBMF	The Shuffled BMF graph dissimilarity implements the same BMF matching scheme. The algorithm allows the user to select a number s , which is the number of shuffles to apply to the vertex set of the input graphs. Consequently, the sBMF algorithm consists in executing the BMF algorithm s times, returning the lower overall dissimilarity value.
	TWEC	The Triple Weighted Error Correcting (TWEC) scheme is basically the normalized Best Match First algorithm, yielding values within the $[0, 1]$ range.
	HGED	This GED is based on the Hungarian algorithm for optimally solving the assignment problem of the vertices. The procedure returns dissimilarity values within the $[0, 1]$ range.
Similarity	GraphCoverage	The Graph Coverage similarity algorithm between arbitrarily labeled graphs. The algorithm was presented here [23, 26], and it is based on the tensor product computation.
MatrixRepresentation	Adjacency	The standard adjacency matrix representation of a graph.
	Eigenvectors	The eigenvectors-based matrix representation for graphs. This class uses the Eigen library.
	InducedTransition	The induced transition matrix is a standard transition matrix computed using the Euclidean norm of the numeric vectors of the edges of the graph.
	Laplacian	The Laplacian matrix representation of a graph.
	NormalizedLaplacian	The normalized Laplacian matrix representation of a graph.
	SymTransition	The symmetric transition matrix representation of a graph.
	Transition	The standard transition matrix representation of a graph.
	WTransition	The transition matrix of a graph using the real-valued labels (the weights) of the graph's edges to induce the transition probabilities.
	HeatMatrix	The heat matrix is elaborated from the normalized graph Laplacian via the exponentiation of the eigenvalues at a given time t .
Operator	TensorProduct	The standard tensor product (also called Kroneker product or direct product) operator between two labeled graphs. The algorithm yields an output graph described in terms of weighted adjacency matrix.
	SymTensorProduct	The tensor product computed assuming symmetric matrices representations (that is, undirected graphs).
	MCTensorProduct	The multicore implementation of the standard tensor product operator.
Seriation	EigenSeriation	The Eigenvectors-based seriation algorithm implemented according to [43]. The algorithm produces an output sequence containing an ordering of the graph vertex labels according to the first eigenvector information of the graph matrix representation.

3 Building Pattern Recognition Applications with SPARE

In this section, we first introduce the idea of generalizing pattern recognition mechanisms. Successively, we consolidate this aspect by providing two different specific application instances. In particular, Section 3.1 discusses, in terms of clustering, two important concepts in this scenario: the dissimilarity measure and a set (cluster) representation technique in the non-geometric case. In Section 3.2 we show the first application involving clustering of real-valued multidimensional sequences, while Section 3.3 shows a benchmark of some graph matching algorithms in the context of classification. In both sections, we report some (minimal) technical details concerning how to combine some of the key SPARE classes needed in those tests. We refer the interested reader to the technical documentation and the practical examples provided within the SPARE installation package for further and more specific details.

3.1 Clustering of a Set of Generic Objects

Cluster analysis, or clustering, is one of the classic problems in pattern recognition. The problem focuses on the task of grouping a set of objects in such a way that objects in the same group (called cluster) are more *similar*, or *compatible*, to each other with respect to those in other clusters. In spite of the availability of a large number of clustering algorithms, and their proved success in a number of different application domains, clustering remains a difficult and ill-posed problem [18]. This fact can be attributed to the inherent vagueness in the definition of a cluster, which magnifies the difficulty of defining an appropriate similarity measure and an effective optimization procedure to guide the clustering process.

The problem of grouping a set of elements into a predefined number k of different clusters is called *k clustering problem*. Given a finite input set $\mathcal{X} = \{x_1, \dots, x_n\}$, where $x_i \in \mathbb{R}^n, i = 1, \dots, n$, a k clustering problem consists in finding a *partition* $\mathcal{P} = \{C_1, \dots, C_k\}$ of this set in a number $k \leq n$ of clusters, with $\mathcal{X} = \bigcup_{i=1}^k C_i$ and $C_i \cap C_j = \emptyset, \forall i, j, i \neq j$, so that for a given metric in \mathbb{R}^n , “close” patterns will fall in the same cluster, while “sufficiently far” patterns will belong to different clusters. In order to solve this problem, the celebrated k -means algorithm [48] tries to minimize the within-cluster sum of squares

$$\arg \min_{\mathcal{P}} \sum_{i=1}^k \sum_{x_j \in C_i} \|x_j - \mu_i\|^2, \quad (1)$$

where μ_i is the *mean* vector of the i -th cluster C_i . The number k is called the *order* of the partition and it is defined a priori. k -means operates as a two-stage iterative *heuristic* algorithm that is based on the assignment of each input pattern x_i to the cluster with the closest mean μ_j . After the *assignment* phase follows the *update* phase, consisting in updating all clusters’ means μ_j . The two phases are iterated until a predefined stop criterion is true.

In order to define clustering algorithms not necessarily dealing with a metric space (e.g., an Euclidean space), we observe some general (*meta*) elements which play a fundamental role in the clustering process. Each cluster is modeled by a *representative*, i.e., an element not necessarily belonging to the input data space, which is able to characterize the entire set of patterns in the cluster. Another important concept is the *dissimilarity measure* used to compute the distance between patterns and representative elements. It is easy to understand that these two definitions of representative of a cluster and of dissimilarity measure strictly depend on the data type characterizing the problem at hand. For example, if $\mathcal{X} = \mathcal{G}$, where \mathcal{G} is a set of graphs, the problem of deriving a representative graph is known as the *median graph* identification [3, 14, 19]. By abstracting the input domain, it is possible to define a *generic* k -means algorithm that uses a generic representative mechanism, which is in turn configured with a generic dissimilarity measure. That is, it is possible to abstract the algorithm itself obtaining thus a *meta-algorithm* that will work on many different problems, but preserving the main behavior (this is the very core of template metaprogramming). The same generalization could be applied to many pattern recognition problems, such as classification and function approximation.

Algorithm 1 shows the pseudo-code of the generic k -means algorithm over an input set $\mathcal{X} \subset \mathcal{U}$ (\mathcal{U} is a given input domain), which is configured to operate by means of k set representatives $\{\mu_1, \dots, \mu_k\}$. The distance

among two input elements is calculated through the dissimilarity measure $d : \mathcal{U} \times \mathcal{U} \rightarrow \mathbb{R}^+$. The pseudo-code of Algorithm 1 shows that the dissimilarity evaluation between a sample x_j and a cluster \mathcal{C}_i can be computed by comparing the sample with the representative of the cluster, namely μ_i , using the dissimilarity function $d(\cdot, \cdot)$ (line 6), defined over the generic space \mathcal{U} . The update mechanism of each representative is carried out by a generic operator $R(\cdot)$ (line 8), able to derive the representative of the modified cluster \mathcal{C}_i . At each step t , a new partition \mathcal{P}^t is induced (line 9), until the stop condition is reached (line 10), which is usually defined as a combination of a maximum number of iterations and other criteria.

Algorithm 1 Generic k -means.

Input: A generic finite input set $\mathcal{X} = \{x_1, \dots, x_n\}$, the order k of the clustering, a dissimilarity function $d : \mathcal{U} \times \mathcal{U} \rightarrow \mathbb{R}_0^+$, the cluster representatives $\{\mu_1, \dots, \mu_k\}$
Output: A partition $\mathcal{P} = \{\mathcal{C}_1, \dots, \mathcal{C}_k\}$
1: Initialize every representative $\mu_i, i = 1, \dots, k$
2: $t = 0, \mathcal{P} = \emptyset$
3: **loop**
4: $t+ = 1$
5: Assignment Step: assign each sample x_j to the cluster with the closest representative
6: $\mathcal{C}_i^{(t)} = \{x_j : d(x_j, \mu_i) \leq d(x_j, \mu_h) \text{ for all } h = 1, \dots, k\}$
7: Update Step: update the representatives
8: $\mu_i^{(t+1)} = R(\mathcal{C}_i), i = 1 \rightarrow k$
9: Update the partition with the modified clusters:
 $\mathcal{P}^t = \{\mathcal{C}_1, \dots, \mathcal{C}_k\}$
10: **if** STOP **then**
11: **return** \mathcal{P}^t
12: **end if**
13: **end loop**

3.1.1 The Dissimilarity Approach in Pattern Recognition

Pattern recognition mechanisms are deeply grounded on the possibility of *distinguishing* among the input patterns, in either a generative or discriminative setting [24, 35]. By departing from the common Euclidean space, we therefore face problems related to the difficulty of defining an effective dissimilarity measure $d : \mathcal{U} \times \mathcal{U} \rightarrow \mathbb{R}^+$ on the input space \mathcal{U} . A dissimilarity is basically the generalization of a metric distance, which may satisfy fewer requirements. When the input space has no trivial geometry, such as in the case of labeled graphs [24] and sequences [29, 41, 42], the definition of such a dissimilarity measure can become a crucial task. Reasonable dissimilarity measures are usually positive and definite functions; however most of the times such algorithms are also symmetric. The triangle inequality instead is not easily achievable in the general case. As pointed out by many authors (see e.g. [24, 30, 35, 36]) from the pattern recognition viewpoint, adhering to the metric requirements does not necessarily imply better performances in terms of pattern recognition problems.

Closely related to the concept of dissimilarity is the one of similarity, which however must be interpreted as the dual concept (it has the dual meaning). The two concepts are intimately related. In fact, it is always possible to derive a similarity measure from the corresponding dissimilarity, and vice versa. A particularly interesting type of similarity measures is provided by *positive definite* kernel functions. A positive definite (pd) kernel function is a symmetric function of the type $k : \mathcal{U} \times \mathcal{U} \rightarrow \mathbb{R}^+$. Well-known pd kernels in pattern recognition are the RBF Gaussian, the Laplacian, the polynomial, and linear kernel [45].

Dissimilarity and pd kernel functions play a pivotal role in implementing generic pattern recognition applications that can operate on user-defined input spaces. SPARE provides the implementation of different dissimilarity measures operating on vectors of real values, sequences of generic objects, labeled graphs, and set of generic objects. Moreover, it includes the implementation of many pd kernel functions. A particular type of kernel functions is the one that is defined in terms of a dissimilarity (such as the RBF Gaussian). In this case, pd kernels are easily extended to different input domains by means of the interoperation between the dissimilarity and similarity concepts.

3.1.2 The MinSOD Cluster Representative

Modeling a finite input set $\mathcal{X} = \{x_1, \dots, x_n\}$ of generic objects is an important aspect of pattern recognition applications. For example, if $\mathcal{X} \subset \mathbb{R}^n$ the representative of such a set may be simply the *centroid*, that is, a vector $\mathbf{x}^* = \frac{1}{n} \sum_i \mathbf{x}_i$, with $\mathbf{x}^* \in \mathbb{R}^n$, which does not necessarily belong to the input set \mathcal{X} . The problem of modeling a set of objects can be generalized to virtually any input domain where it is possible to define a dissimilarity measure $d(\cdot, \cdot)$. Formally, the representative element can be computed [12] as follows:

$$x^* = \arg \min_{x_i \in \mathcal{X}} \sum_{x_j \in \mathcal{X}, x_j \neq x_i} d(x_j, x_i). \quad (2)$$

The *set mean* in this case is taken as the element of the set \mathcal{X} that minimizes the sum of the distances (MinSOD) between the element itself and the other elements in the set. To obtain the MinSOD we need to execute a quadratic number of distance calculations. That is, the overall complexity depends in turn on the computational complexity of $d(\cdot, \cdot)$, which, especially when dealing with complex data structures [24], is usually polynomial. If \mathcal{X} is a large set, the evaluation of the distance between every objects of \mathcal{X} become prohibitive and some approximation mechanism must be taken into account.

In SPARE the MinSOD determination is achieved through a specific class, named *MinSod*, which models the *Representative* concept. The *MinSod* object performs an important computational speed up by only determining the MinSOD representative inside a reduced pool of samples, instead of the whole set of the input samples inserted in the cluster so far. The reduced pool of samples is called the *cache* and its size is a relevant parameter for the user to setup. In order to work with a reduced set of samples, a *replacement policy* has to be defined to discard some samples from the pool as new samples are inserted in the set. In SPARE, the MinSOD has been implemented according to different replacement policies, which are specialized for different purposes. The standard MinSOD replacement policy is simple but efficient in practice. Each new element is inserted since the defined cache size is reached. Then, for each new sample an old one must be discarded. The old sample to discard is chosen as follows: two samples are chosen with uniform probability, and the one farthest away from the actual MinSOD representative is discarded. We refer the interested reader to the SPARE technical documentation for more details. In Algorithm 2 is shown the pseudo-code of the behavior of the *MinSod* class of SPARE.

Algorithm 2 Standard MinSOD representative and related cache replacement.

Input: A generic finite input set $\mathcal{X} = \{x_1, \dots, x_n\}$, the size M of the cache \mathcal{C} , a dissimilarity function $d : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}_0^+$

Output: The MinSOD determination $\hat{x} \in \mathcal{X}$

```

1:  $\mathcal{C} = \emptyset$ 
2: for all  $x_i \in \mathcal{X}$  do
3:   if  $|\mathcal{C}| < M$  then
4:      $\mathcal{C} = \mathcal{C} \cup \{x_i\}$ 
5:   else
6:     Select two elements, namely  $\chi_1$  and  $\chi_2$  from  $\mathcal{C}$  with uniform probability
7:     if  $d(\chi_1, \hat{x}) \geq d(\chi_2, \hat{x})$  then
8:        $\mathcal{C} = \mathcal{C} \setminus \{\chi_1\}$ 
9:     else
10:       $\mathcal{C} = \mathcal{C} \setminus \{\chi_2\}$ 
11:    end if
12:     $\mathcal{C} = \mathcal{C} \cup \{x_i\}$ 
13:  end if
14: Determine  $\hat{x} = \text{MinSOD}(\mathcal{C})$  according to Equation 2
15: end for
```

3.2 Sequence Clustering and MinSOD Representative Evaluation

In this section we deal with an input set containing real-valued multidimensional sequences. In particular, we will conceive a clustering problem involving the aforementioned standard MinSOD cluster model, testing its robustness when varying the cache size parameter. Since we deal with input patterns that are sequences of

real-valued vectors, we will equip the MinSOD representative with a dissimilarity measure elaborated from the well-known Dynamic Time Warping (DTW) matching algorithm [7, 17, 29, 33, 41, 42, 44]. The DTW follows a global alignment scheme, which has been proposed to align sequences of arbitrarily-defined objects belonging to a set \mathcal{A} (e.g., sequences of characters, sequences of real-valued vectors, or even sequences of complex data structures). The key component of the algorithm is the core dissimilarity measure, $d_{\mathcal{A}} : \mathcal{A} \times \mathcal{A} \rightarrow \mathbb{R}^+$, which quantifies the cost of comparing two input objects. In the herein discussed problem, the objects are 5-dimensional real-valued vectors, and the core dissimilarity measure of the DTW is the Euclidean metric.

The considered clustering problem is defined as a two-class classification problem, with each class being modeled by a single cluster and where each object of the sequence is a 5-dimensional real-valued vector extracted from a 5-dimensional Gaussian distribution. Each class contain 150 sequences (for a total of 300 sequences for each problem instance), and it is characterized by a specific Gaussian distribution with a certain mean and (spherical) covariance matrix. By fixing the mean value of the two 5-dimensional Gaussian distributions, we can control the class overlap by increasing the variance parameter of the respective covariance matrices. The objective is to group sequences that pertain to the same generating class, that is, are generated from the same Gaussian distribution. This is in all respects an unsupervised learning task. Once the partition is obtained, a simple quality index which assumes values in $[0, 1]$ is evaluated for each computed cluster. The quality index takes into account the *purity* of the cluster in terms of the known class labels; it is defined by the number of samples belonging to the majority class in the cluster, divided by the total cardinality of the cluster. An index value of one indicates a pure cluster, composed by samples of the same class only. An index value of 0.5 indicates a cluster which contains half of the samples from one class, half from the other. The overall quality index is taken as the worst quality index over the two clusters. We show the results obtained over two different tests, with the second problem representing a more difficult clustering problem (higher overlap of the Gaussian distributions) with respect to the first one. Our aim is to observe how, and if, reducing the cache size of the MinSOD affects the established performance measure. To this end, the cache size is decremented, for both tests, starting from a value of 50 and going down to 1, with a unit decrement step. For each size of the cache, the same test is carried out with five different random seeds. The results for the quality index are eventually taken as the average over these runs; the worst and better case results are also considered at each cache size.

3.2.1 Code Snippets

The practical SPARE configuration of the DTW template class for processing such input patterns looks like:

Listing 1: DTW configuration for processing sequences of real-valued vectors.

```
typedef std::vector<double> SequenceObjectType;
typedef std::vector<SequenceObjectType> SequenceType;
typedef Euclidean SequenceObjectDissimilarityType;
typedef Dtw<SequenceObjectDissimilarityType> SequenceDissimilarityType;
```

Once we have defined the dissimilarity measure for the considered type of sequences, we also need to define the MinSOD cluster model type, and finally the clustering algorithm, which operates by using such cluster models. We adopt the k -means implementation of SPARE as the clustering algorithm. The SPARE code to set up such classes via type definitions is:

Listing 2: Configuration of the MinSOD and clustering algorithm.

```
typedef MinSod<SequenceType, SequenceDissimilarityType> MinSodType;
typedef FirstK KmeansInitializerType;
typedef Kmeans<MinSodType, KmeansInitializerType> KmeansAlgoType;
```

The herein prepared k -means type is then instantiated as a runtime object and used via the main method called *Process*, by using the standard STL container-based data passing interface. In Listing 3 the call to the *Process* method takes as the first two arguments the starting and ending iterators delimiting the patterns to be analyzed into the container (assumed to be stored into the *DataSet* container). The third argument,

i.e., the container called *Labels*, accounts for the output labels describing the calculated cluster associations of the input patterns.

Listing 3: Processing of the input data.

```
std::vector<SequenceType> DataSet;
std::vector<NaturalType> Labels;
...
KmeansAlgoType clusteringAlgorithm;
...
clusteringAlgorithm.Process(DataSet.begin(),
                           DataSet.end(), Labels.begin());
```

The user can then use those output labels, or the set of calculated cluster representatives accessible via the clustering algorithm object, for its own practical purposes.

It is very easy to understand that such a setup can be very easily generalized to any user-defined data type. Moreover, switching among the possible cluster models and pattern dissimilarity measures is totally independent from the clustering algorithm functionality, which simplifies the practical development of generic and reusable pattern recognition systems.

3.2.2 Results

The results of first test are shown in Figure 1(a). It is possible to observe a clear breakdown at a cache size of dimension 5. For higher values the results are very stable and nearly optimal (i.e., close to one), meaning that the MinSOD cluster models effectively the input classes. Also in the second test the cache size of MinSOD varies from 50 to 1, with a decrement step of 1; the same considerations about the random seed initialization hold also in this case. The results for this second test are shown in Figure 1(b). In this case, since the problem is more difficult, performances start to seriously deteriorate from a cache size around 10. The performance tests show that the implemented MinSOD algorithm is able to model effectively the input set, with a correct behavior even when reducing the cache size. In fact, before reaching a very low critical threshold for the cache size, the size is quite irrelevant for the cluster modeling purpose. A very good stability is observed with respect to the randomized replacement policy. This stability is observed along the whole interval of the valid settings of the cache size, while strong variations arise only in the region where the cache size is too low to guarantee a correct coverage of the input set.

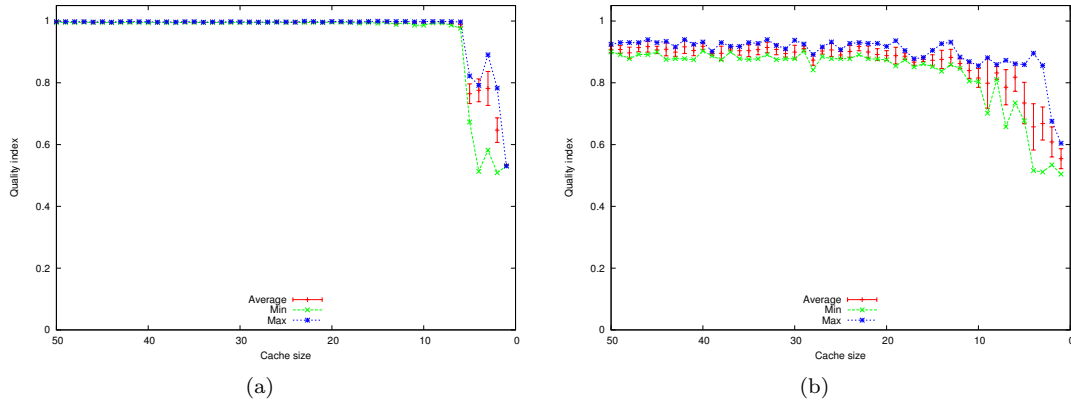


Figure 1: The average performance results (quality index vs cache size) are reported together with the standard deviation (vertical bars). 1(a) shows the results for the first test, while 1(b) for the second one.

3.3 Benchmarking Graph Matching Algorithms in Classification Problems

As a further example of the versatility of the SPARE design, in this section we discuss the application of some graph matching algorithms implemented in SPARE, on different classification problem instances defined on synthetically generated labeled graphs. Graph matching algorithms play a key role in the design of graph-based pattern recognition systems [5, 8, 9, 15, 24, 27, 30, 31, 34], since they provide the basic mechanism to measure the graph dissimilarity value.

Each synthetic dataset has been generated using the same Markov chains based method described by Livi et al. [25]. We have conceived 15 different two-classes classification problem instances, which are characterized by a decreasing level of difficulty. Each problem is defined by a training, a validation, and a test set, containing 300 graphs each of order 30 and a variable size between 40 and 75. The hardness of the problem has been controlled generating the labels of both vertices and edges as numeric vectors falling in $[0, 1]^5$. Each random vector has been constructed sampling numbers from two different Gaussian distributions (i.e., one for each class), distributed with suitably chosen means and standard deviations. In addition, the stochastic generation process contributes in generating graphs with a randomized topology.

In this experiment, we have considered four different graph classification systems, which are based on a classifier relying on the k -NN rule. The k -NN rule offers a perfect example to employ the concept of dissimilarity operating directly on the input space. In fact, the neighborhood selection, as well as the classification rule, is carried out completely on the base of the outcomes yielded by the adopted dissimilarity measure. Therefore, since we deal with labeled graphs, we test three k -NN based graph classification systems that operate directly on the input space \mathcal{G} of labeled graphs by means of suited inexact graph matching algorithms [24]. Two of the considered algorithms adopt the BMF-GED algorithm using the 6 weights (PD6W) and the 3 weights (TWEC) schemes for the overall edit distance computation [24]. The third one instead is a graph kernel function called *Graph Coverage* (GC) [23, 26]. Finally, a graph seriation-based system equipped with the DTW distance (Seriation), tailored for sequences of five-dimensional real-valued vectors, is considered [25].

3.3.1 Code Snippets

In Listing 4, we briefly summarize the main SPARE classes involved in the configuration of such a graph classifier in the TWEC case; other cases follow straightforwardly. The first thing to define is the type of considered labeled graph. Since we deal with graphs having vertex and edge labels defined as 5-dimensional real-valued vectors, the SPARE graph type to be considered is the “VectorGraph” type. Such a graph structure is provided with the SPARE graph package; however, the user can define virtually any type of labeled graph. The other involved types define the graph dissimilarity (e.g., the TWEC graph matching in this example) and the k -NN classifier. Such a classifier is able to operate directly on the considered labeled graph domain by means of the direct evaluations of the graph dissimilarity measure.

Listing 4: k -NN Classifier equipped with the TWEC graph dissimilarity measure.

```
typedef VectorGraph LabeledGraphType;
typedef string LabelType;
typedef Euclidean VertexLabelsDissimilarityType;
typedef Euclidean EdgeLabelsDissimilarityType;
typedef TWEC<VertexLabelsDissimilarityType,
            EdgeLabelsDissimilarityType> GraphDissimilarityType;
typedef KnnClass<LabeledGraphType, GraphDissimilarityType,
                LabelType> ClassifierType;
```

By making use of the data types defined in the previous listing, we can now allocate and configure the graph classifier. After the object allocation, we can train the classifier by using the *Learn* method (note however that the herein considered k -NN does not have a proper model synthesis stage; however this scheme applies to any SPARE classifiers which do also synthesize a classification model) and then test the synthesized model by calling the *Process* method, as shown in Listing 5. The *Learn* method accepts three arguments: the first two are the iterators delimiting the input training graphs into the container, while the third one is

an iterator pointing at the first class label corresponding to the first training graph. The *Process* method works analogously, although the method writes in the container referenced by the output label iterator (third argument) the calculated class labels for the test graphs. The user is then in charge of defining a meaningful performance (i.e., error or accuracy) measure with respect to the effective test graph class labels.

Listing 5: Learning and testing the graph classifier.

```
std::vector<LabeledGraphType> trainGraphs, testGraphs;
std::vector<LabelType> trainLabels, testLabels, outLabels;
...
ClassifierType graphClassifier;
...
graphClassifier.Learn(trainGraphs.begin(), trainGraphs.end(),
                    trainLabels.begin());

//output labels allocation
outLabels.assign(testGraphs.size(), "undetermined");
graphClassifier.Process(testGraphs.begin(), testGraphs.end(),
                      outLabels.begin());
```

Again, it is very easy to understand that the generalization of the k -NN rule on other data types is very straightforward, requiring few practical changes in herein presented setup.

3.3.2 Results

Tests have been performed using three values for the k parameter defining the considered nearest neighbors in the k -NN classification rule: namely 1, 3, and 5. Since the considered graph matching algorithms are sensible to the setting of the weighting parameters, we optimize those parameters by a cross validation procedure. The global optimization is performed by means of the genetic algorithm implemented in SPARE. The fitness function guiding the optimization is thus defined considering directly the recognition rate achieved on the validation set. The setup and integration of the genetic algorithm is omitted here for the sake of brevity; the interested reader can however refer to the “GKnn.cpp” example provided with the SPARE package.

Figures 2(a), 2(b), and 2(c) show the obtained performances (test classification accuracy in our case) on the batch of synthetic tests. Clearly, the system configured with the GC graph matching algorithm achieves the best test set classification accuracy results considering the three values of k . TWEC and PD6W obtain comparable results, while the former achieves in general slightly better accuracy values.

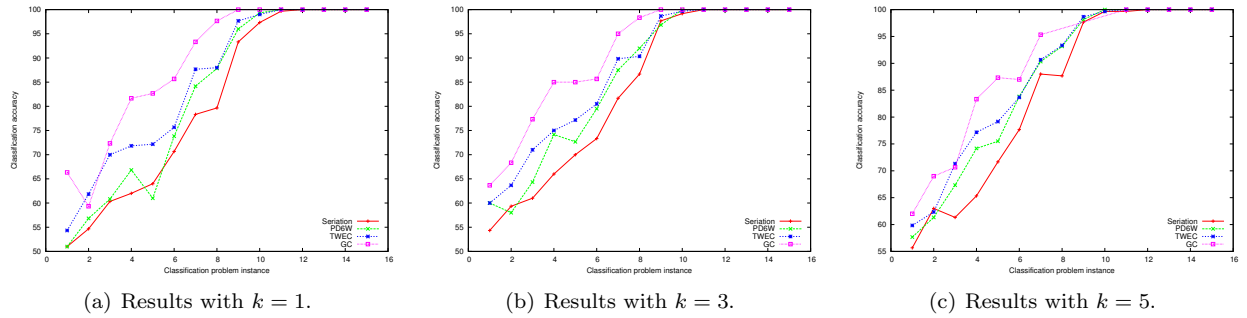


Figure 2: Test set classification accuracy results of the graph-based classification system over the batch of 15 problem instances.

4 Conclusions

In this paper we have presented the SPARE library as an open source tool to rapidly develop generic pattern recognition applications. We introduced the underlying design of the library, which is based on the template metaprogramming framework offered by the C++ programming language. Main functionalities provided by SPARE have been briefly highlighted. As practical examples of the SPARE features and versatility, we have shown two applications involving clustering and classification problems defined on non-geometric input spaces. The former has been conceived to show the generic MinSOD cluster representation, applied to the context of clustering of real-valued multidimensional sequences. As a further example, we discussed also the classification of labeled graphs by means of the k -NN rule, operating directly on the input space of labeled graphs. All source codes and tutorials can be retrieved together with the latest SPARE release.

Future directions pertain, of course, the development of new algorithms, ranging from new clustering algorithms to additional meta-heuristic global optimization techniques and new graph matching procedures.

Acknowledgements

This work was partially supported by the “POlo per la MObilità Sostenibile” (POMOS), the research center for sustainable mobility, Lazio Region, Italy.

References

- [1] D. Abrahams and A. Gurtovoy. *C++ template metaprogramming: concepts, tools, and techniques from boost and beyond*. C++ in-depth series. Addison-Wesley, 2005. ISBN 9780321227256.
- [2] G. Attardi and A. Cisternino. Reflection Support by Means of Template Metaprogramming. In *Proceedings of the Third International Conference on Generative and Component-Based Software Engineering, GCSE '01*, pages 118–127, London, UK, 2001. Springer-Verlag. ISBN 3-540-42546-2.
- [3] I. Bardaji, M. Ferrer, and A. Sanfeliu. A comparison between two representatives of a set of graphs: median vs. barycenter graph. In *Proceedings of the 2010 joint IAPR international conference on Structural, syntactic, and statistical pattern recognition, SSPR&SPR'10*, pages 149–158, Berlin, Heidelberg, 2010. Springer-Verlag. ISBN 3-642-14979-0, 978-3-642-14979-5.
- [4] T. Becker. STL Containers. *C/C++ Users Journal*, 19:62–70, Feb. 2001. ISSN 1075-2838.
- [5] E. Bengoetxea, P. Larrañaga, I. Bloch, A. Perchant, and C. Boeres. Inexact graph matching by means of estimation of distribution algorithms. *Pattern Recognition*, 35(12):2867–2880, 2002. ISSN 0031-3203. doi: 10.1016/S0031-3203(01)00232-1.
- [6] M. Benzinger. STL containers based on hash tables. *C/C++ Users Journal*, 16:43–47, Feb. 1998. ISSN 1075-2838.
- [7] D. J. Berndt and J. Clifford. Advances in knowledge discovery and data mining. chapter Finding patterns in time series: a dynamic programming approach, pages 229–248. American Association for Artificial Intelligence, Menlo Park, CA, USA, 1996. ISBN 0-262-56097-6.
- [8] F. M. Bianchi, L. Livi, and A. Rizzi. Matching of time-varying labeled graphs. In *Proceedings of the 2013 International Joint Conference on Neural Networks*, pages 1660–1667, aug 2013. ISBN 978-1-4673-6129-3. doi: 10.1109/IJCNN.2013.6706939.
- [9] F. M. Bianchi, L. Livi, A. Rizzi, and A. Sadeghian. A Granular Computing approach to the design of optimized graph classification systems. *Soft Computing*, 18(2):393–412, 2014. ISSN 1432-7643. doi: 10.1007/s00500-013-1065-z.
- [10] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3):27:1–27:27, May 2011. ISSN 2157-6904. doi: 10.1145/1961189.1961199.

- [11] G. Del Vescovo and A. Rizzi. Online Handwriting Recognition by the Symbolic Histograms Approach. In *Proceedings of the 2007 IEEE International Conference on Granular Computing*, GRC '07, pages 686–700, Washington, DC, USA, 2007. IEEE Computer Society. ISBN 0-7695-3032-X. doi: 10.1109/GRC.2007.116.
- [12] G. Del Vescovo, L. Livi, F. M. Frattale Mascioli, and A. Rizzi. On the problem of modeling structured data with the MinSOD representative. *International Journal of Computer Theory and Engineering*, 6(1):9–14, 2014. ISSN 1793-8201. doi: 10.7763/IJCTE.2014.V6.827.
- [13] G. Dos Reis and B. Stroustrup. Specifying C++ concepts. *SIGPLAN Not.*, 41:295–308, January 2006. ISSN 0362-1340. doi: 10.1145/1111320.1111064.
- [14] M. Ferrer, E. Valveny, F. Serratos, K. Riesen, and H. Bunke. Generalized median graph computation by means of graph embedding in vector spaces. *Pattern Recogn.*, 43:1642–1655, Apr. 2010. ISSN 0031-3203. doi: 10.1016/j.patcog.2009.10.013.
- [15] X. Gao, B. Xiao, D. Tao, and X. Li. A survey of graph edit distance. *Pattern Analysis and Applications*, 13(1): 113–129, Feb. 2010. ISSN 1433-7541. doi: 10.1007/s10044-008-0141-y.
- [16] A. Giuliani, R. Benigni, J. P. Zbilut, C. L. Webber Jr., P. Sirabella, and A. Colosimo. Nonlinear Signal Analysis Methods in the Elucidation of Protein Sequence—Structure Relationships. *ChemInform*, 33(28):1471–1492, 2002. ISSN 1522-2667. doi: 10.1002/chin.200228300.
- [17] H. Izakian, W. Pedrycz, and I. Jamal. Fuzzy clustering of time series data using dynamic time warping distance. *Engineering Applications of Artificial Intelligence*, 39:235–244, 2015. doi: 10.1016/j.engappai.2014.12.015.
- [18] A. K. Jain. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8):651–666, June 2010. ISSN 0167-8655. doi: 10.1016/j.patrec.2009.09.011.
- [19] X. Jiang, A. Münger, and H. Bunke. On Median Graphs: Properties, Algorithms, and Applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:1144–1151, October 2001. ISSN 0162-8828. doi: 10.1109/34.954604.
- [20] S. C. Johnson. YACC—yet another compiler-compiler. Technical Report CS-32, AT & T Bell Laboratories, Murray Hill, N.J., 1975.
- [21] B. Karlsson. *Beyond the C++ standard library: an introduction to Boost*. Addison-Wesley, 2006. ISBN 9780321133540.
- [22] D. E. King. Dlib-ml: A Machine Learning Toolkit. *Journal of Machine Learning Research*, 10:1755–1758, 2009.
- [23] L. Livi and A. Rizzi. Parallel algorithms for tensor product-based inexact graph matching. In *Proceedings of the 2012 International Joint Conference on Neural Networks*, pages 2276–2283, June 2012. ISBN 978-1-4673-1489-3. doi: 10.1109/IJCNN.2012.6252681.
- [24] L. Livi and A. Rizzi. The graph matching problem. *Pattern Analysis and Applications*, 16(3):253–283, 2013. ISSN 1433-7541. doi: 10.1007/s10044-012-0284-8.
- [25] L. Livi, G. Del Vescovo, and A. Rizzi. Graph recognition by seriation and frequent substructures mining. In *Proceedings of the First International Conference on Pattern Recognition Applications and Methods*, volume 1, pages 186–191, Feb. 2012. ISBN 978-989-8425-98-0. doi: 10.5220/0003733201860191.
- [26] L. Livi, G. Del Vescovo, and A. Rizzi. Inexact graph matching through graph coverage. In *Proceedings of the First International Conference on Pattern Recognition Applications and Methods*, volume 1, pages 269–272, Feb. 2012. ISBN 978-989-8425-98-0. doi: 10.5220/0003732802690272.
- [27] L. Livi, F. M. Bianchi, A. Rizzi, and A. Sadeghian. Dissimilarity space embedding of labeled graphs by a clustering-based compression procedure. In *Proceedings of the 2013 International Joint Conference on Neural Networks*, pages 1646–1653, Aug 2013. ISBN 978-1-4673-6129-3. doi: 10.1109/IJCNN.2013.6706937.

- [28] L. Livi, G. Del Vescovo, and A. Rizzi. Combining graph seriation and substructures mining for graph recognition. In P. Latorre Carmona, J. S. Sánchez, and A. L. N. Fred, editors, *Pattern Recognition - Applications and Methods*, volume 204 of *Advances in Intelligent and Soft Computing*, pages 79–91. Springer Berlin Heidelberg, 2013. ISBN 978-3-642-36529-4. doi: 10.1007/978-3-642-36530-0_7.
- [29] L. Livi, H. Tahayori, A. Sadeghian, and A. Rizzi. Aggregating α -planes for type-2 fuzzy set matching. In *2013 Joint IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS)*, pages 860–865, 2013. doi: 10.1109/IFSA-NAFIPS.2013.6608513.
- [30] L. Livi, A. Rizzi, and A. Sadeghian. Optimized dissimilarity space embedding for labeled graphs. *Information Sciences*, 266:47–64, 2014. ISSN 0020-0255. doi: 10.1016/j.ins.2014.01.005.
- [31] L. Livi, A. Rizzi, and A. Sadeghian. Designing labeled graph classifiers by exploiting the Rényi entropy of the dissimilarity representation. *ArXiv preprint arXiv:1408.5286*, Aug 2014.
- [32] B. Long. Towards the design of a set-based Java collections framework. *SIGSOFT Softw. Eng. Notes*, 35:1–7, October 2010. ISSN 0163-5948. doi: 10.1145/1838687.1838705.
- [33] T. Nakamura, K. Taki, H. Nomiya, K. Seki, and K. Uehara. A shape-based similarity measure for time series data with ensemble learning. *Pattern Analysis & Applications*, pages 1–14, 2011. ISSN 1433-7541. 10.1007/s10044-011-0262-6.
- [34] A. Noma, A. B. V. Graciano, R. M. Cesar Jr, L. A. Consularo, and I. Bloch. Interactive image segmentation by matching attributed relational graphs. *Pattern Recognition*, 45(3):1159–1179, 2012. ISSN 0031-3203. doi: 10.1016/j.patcog.2011.08.017.
- [35] E. Pękalska and R. P. W. Duin. *The dissimilarity representation for pattern recognition: foundations and applications*. Series in machine perception and artificial intelligence. World Scientific, 2005. ISBN 9789812565303.
- [36] E. Pękalska, A. Harol, R. P. W. Duin, B. Spillmann, and H. Bunke. Non-Euclidean or Non-metric Measures Can Be Informative. In D.-Y. Yeung, J. Kwok, A. L. N. Fred, F. Roli, and D. Ridder, editors, *Structural, Syntactic, and Statistical Pattern Recognition*, volume 4109 of *LNCIS*, pages 871–880. Springer Berlin Heidelberg, 2006. ISBN 978-3-540-37236-3. doi: 10.1007/11815921_96.
- [37] S. Queiroz, F. de Carvalho, and Y. Lechevallier. Nonlinear multicriteria clustering based on multiple dissimilarity matrices. *Pattern Recognition*, 46(12):3383 – 3394, 2013. ISSN 0031-3203. doi: 10.1016/j.patcog.2013.06.008.
- [38] K. Riesen and H. Bunke. *Graph Classification and Clustering Based on Vector Space Embedding*. Series in Machine Perception and Artificial Intelligence. World Scientific Pub Co Inc, 2010. ISBN 9789814304719.
- [39] A. Rizzi, M. Panella, and F. M. Frattale Mascioli. Adaptive resolution min-max classifiers. *IEEE Transactions on Neural Networks*, 13:402–414, Mar. 2002. ISSN 1045-9227.
- [40] A. Rizzi, G. Del Vescovo, L. Livi, and F. M. Frattale Mascioli. A new granular computing approach for sequences representation and classification. In *Proceedings of the 2012 International Joint Conference on Neural Networks*, pages 2268–2275, June 2012. ISBN 978-1-4673-1489-3. doi: 10.1109/IJCNN.2012.6252680.
- [41] A. Rizzi, L. Livi, H. Tahayori, and A. Sadeghian. Matching general type-2 fuzzy sets by comparing the vertical slices. In *2013 Joint IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS)*, pages 866–871, 2013. doi: 10.1109/IFSA-NAFIPS.2013.6608514.
- [42] A. Rizzi, F. Possemato, L. Livi, A. Sebastiani, A. Giuliani, and F. M. Frattale Mascioli. A dissimilarity-based classifier for generalized sequences by a Granular Computing approach. In *Proceedings of the 2013 International Joint Conference on Neural Networks*, pages 2397–2404, Aug 2013. ISBN 978-1-4673-6129-3. doi: 10.1109/IJCNN.2013.6707041.
- [43] A. Robles-Kelly and E. R. Hancock. Graph Edit Distance from Spectral Seriation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:365–378, March 2005. ISSN 0162-8828. doi: 10.1109/TPAMI.2005.56.
- [44] H. Sakoe. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26:43–49, 1978.

- [45] B. Schölkopf and A. J. Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. Adaptive computation and machine learning. MIT Press, 2002. ISBN 9780262194754.
- [46] B. Stroustrup. *The C++ Programming Language*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 3rd edition, 2000. ISBN 0201700735.
- [47] B. Stroustrup. Evolving a language in and for the real world: C++ 1991-2006. In *Proceedings of the third ACM SIGPLAN conference on History of programming languages*, HOPL III, pages 4–1–4–59, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-766-7. doi: 10.1145/1238844.1238848.
- [48] S. Theodoridis and K. Koutroumbas. *Pattern recognition*. Elsevier/Academic Press, 4th edition, 2008. ISBN 9780123695314.
- [49] K. Tun, P. Dhar, M. Palumbo, and A. Giuliani. Metabolic pathways variability and sequence/networks comparisons. *BMC Bioinformatics*, 7(1):24, 2006. ISSN 1471-2105. doi: 10.1186/1471-2105-7-24.
- [50] D. Vandevoorde and N. Josuttis. *C++ templates: the complete guide*. Addison-Wesley, 2003. ISBN 9780201734843.
- [51] T. Veldhuizen. *Using C++ template metaprograms*, pages 459–473. SIGS Publications, Inc., New York, NY, USA, 1996. ISBN 1-884842-37-2.